CRM 90-182 / December 1990

Validation of ASVAB Against Infantry Job Performance

Paul W. Mayberry

CNA
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

Work conducted under contract N00014-91-C-0002.

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1990	Final

4. TITLE AND SUBTITLE
Validation of ASVAB Against Infantry Job Performance

5. FUNDING NUMBERS
C - N00014-91-C-0002
PE - 65153M
PR - C0031

6. AUTHOR(S)
Paul W. Mayberry

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Naval Analyses
4401 Ford Avenue
Alexandria, Virginia 22302-0268

8. PERFORMING ORGANIZATION REPORT NUMBER
CRM 90-182

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commanding General
Marine Corps Combat Development Command (WF 13F)
Studies and Analyses Branch
Quantico, Virginia 22134

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. *SUPPLEMENTARY NOTES*

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This research memorandum examines the validity of the Armed Services Vocational Aptitude Battery (ASVAB) against several measures of infantry performance collected as part of the Job Performance Measurement (JPM) project. ASVAB was highly related to hands-on performance of critical infantry tasks, written infantry job knowledge tests, and grades in initial infantry training courses. The validity of ASVAB was also found to extend into the second term of service by being able to predict the performance of infantry unit leaders.

Experience in the Marine Corps, as measured by time in service, rank, and recency of task performance, had a significant and positive influence on infantry performance. Despite such developmental effects, experience was not able to overcome the significant performance differentials found between low- and high-aptitude perfonnel.

The Marine Corps JPM project succeeded in developing reliable, objective, and representative measures of infantry performance. The results of the project provide the Marine Corps with convincing empirical evidence that demonstrates the strong linkage between performance on the ASVAB abd future infantry performance.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Aptitude tests, Aptitudes, ASVAB (armed services vocational aptitude battery), Demography, Infantry personnel, Job training, Marine corps personnel, Marine corps training, Performance (human), Performance tests, Predictions, Proficiency, Statistical analysis, Validation	57
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
CPR	CPR	CPR	SAR

CNA CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268 • (703) 824-2000

17 January 1991

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Memorandum 90-182

Encl: (1) CNA Research Memorandum 90-182, *Validation of ASVAB Against Infantry Job Performance*, by Paul W. Mayberry, Dec 1990

1. Enclosure (1) is forwarded as a matter of possible interest.

2. This research memorandum examines the validity of the Armed Services Vocational Aptitude Battery (ASVAB) against several measures of infantry performance collected as part of the Job Performance Measurement (JPM) project. ASVAB was found to be highly related to hands-on performance of critical infantry tasks, written infantry job knowledge tests, and grades in initial infantry training courses. The validity of ASVAB was also found to extend into the second term of service by being able to predict the performance of infantry unit leaders.

3. Experience in the Marine Corps, as measured by time in service, rank, and recency of task performance, had a significant and positive influence on infantry performance. Despite such developmental effects, experience was not able to overcome the significant performance differentials found between low- and high-aptitude personnel.

4. The Marine Corps JPM project succeeded in developing reliable, objective, and representative measures of infantry performance. The results of the project provide the Marine Corps with convincing empirical evidence that demonstrates the strong linkage between performance on the ASVAB and future infantry performance.

Lewis R. Cabe
Director
Manpower and Training Program

Distribution List:
Reverse page

Subj: Center for Naval Analyses Research Memorandum 90-182

Distribution List

SNDL		OPNAV
24J1	CG FMFLANT	OP-01
24J2	CG FMFPAC	OP-11
45A2	CG I MEF	OP-12
45A2	CG II MEF	OP-13
45A2	CG III MEF	
45B	CG FIRST MARDIV	
45B	CG SECOND MARDIV	
A1	ASSTSECNAV MRA	
A1	DASN MANPOWER (2 copies)	
A2A	CNR	
A6	HQMC MPR & RA	
	Attn: Code M	
	Attn: Code MR	
	Attn: Code MP	
	Attn: Code MM	
	Attn: Code MA (3 copies)	
	Attn: Code MPP-54	
A6	HQMC PP&O	
A6	CG MCRDAC, Washington	
A6	HQMC AVN	
FF38	USNA	
	Attn: Nimitz Library	
FF42	NAVPGSCOL	
FF44	NAVWARCOL	
FJA1	COMNAVMILPERSCOM	
FJB1	COMNAVCRUITCOM	
FJA13	NAVPERSRANDCEN	
	Attn: Technical Director (Code 01)	
	Attn: Technical Library	
	Attn: Director, Manpower Systems (Code 11)	
	Attn: Director, Personnel Systems (Code 12)	
	Attn: Director, Testing Systems (Code 13)	
	Attn: CAT/ASVAB PMO	
FT1	CNET	
V8	CG MCRD Parris Island	
V8	CG MCRD San Diego	
V12	MCCDC	
	Attn: Studies and Analyses Branch	
	Attn: Director, Warfighting Center	
	Attn: Warfighting Center, MAGTF Proponency and Requirements Branch (2 copies)	
	Attn: Director, Training and Education Center	
V12	CG MCRDAC, Quantico	

CRM 90-182 / December 1990

Validation of ASVAB Against Infantry Job Performance

Paul W. Mayberry

Force Structure and Acquisition Division

CNA
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

ABSTRACT

This research memorandum examines the validity of the Armed Services Vocational Aptitude Battery (ASVAB) against several measures of infantry performance collected as part of the Job Performance Measurement (JPM) project. ASVAB was highly related to hands-on performance of critical infantry tasks, written infantry job knowledge tests, and grades in initial infantry training courses. The validity of ASVAB was also found to extend into the second term of service by being able to predict the performance of infantry unit leaders.

Experience in the Marine Corps, as measured by time in service, rank, and recency of task performance, had a significant and positive influence on infantry performance. Despite such developmental effects, experience was not able to overcome the significant performance differentials found between low- and high-aptitude personnel.

The Marine Corps JPM project succeeded in developing reliable, objective, and representative measures of infantry performance. The results of the project provide the Marine Corps with convincing empirical evidence that demonstrates the strong linkage between performance on the ASVAB and future infantry performance.

EXECUTIVE SUMMARY

The joint-service Job Performance Measurement (JPM) project was initiated in 1981 to link military aptitude requirements to objective measures of job performance. Early in the project's history, hands-on performance tests (HOPTs) were established as the benchmark measure of job performance. HOPTs have high fidelity to actual job performance[1] and require the examinee to perform critical job tasks under realistic but standardized testing conditions. This research memorandum investigates the relationship between these hands-on performance measures and the test used for selection into the military, the Armed Services Vocational Aptitude Battery (ASVAB). Implications for a variety of Marine Corps issues related to manpower, recruiting, and training are also noted.

The Marine Corps focused on the infantry occupational field as its contribution to the joint-service project. More than 2,300 infantrymen from five infantry specialties were tested: rifleman, machinegunner, mortarman, assaultman, and infantry unit leader. Each Marine was tested for two days. All examinees were administered hands-on performance tests and job knowledge tests (JKTs). A variety of other performance information was also collected from historical personnel records, which included proficiency ratings (PRO marks) and grade point averages in initial infantry training.

RESULTS

Relationship Between ASVAB and Infantry Performance

The Marine Corps uses the general technical (GT) aptitude score, a composite of ASVAB subtests, to determine eligibility for military occupational specialties (MOSs) in the infantry occupational field. The validity of the GT aptitude composite against HOPTs indicated a strong relationship between an individual's aptitude level and his later performance of critical infantry tasks. This relationship was also found to extend into the second term for the performance of tasks specifically required of infantry unit leaders. Validities of GT aptitude scores against the JKTs were somewhat higher than those for the HOPTs, but validities for

1. The term "job proficiency" is used throughout this research memorandum to refer to performance on the hands-on performance tests. Conversely, "job performance" is a more general term and is used to reflect what individuals do on their jobs or the assessment of individuals' abilities in a generic sense.

proficiency marks were substantially lower. The JKT results were comparable to previous validation research using training grades as the criterion. The low PRO mark results have also been found in other studies and may be due to the subjectivity of supervisor ratings or the lack of PRO mark variability.

These validity findings imply that high-aptitude personnel perform better than low-aptitude personnel, but how much better? Individuals of high aptitude (GT greater than 109) were found to perform about 25 percent better on HOPTs than their counterparts with low aptitudes (GT less than 90). Likewise, substantial JKT differences were noted for the lowest versus highest aptitude groups—over 30 percent. Consistent with the lack of validity for PRO marks, persons with high GT scores received ratings only 2 percent higher than persons with the low GT scores.

Interaction of Aptitude and Job Experience

Performance is expected to improve as Marines receive more experience through on-the-job training and advanced instruction. The interaction of experience and aptitude in predicting performance has significant implications for Marine Corps recruiting: Do high- and low-aptitude individuals differ in their performance across all levels of experience, or can experience compensate for lower aptitude? The aptitude/experience interaction was examined for three measures of experience: time in service, rank, and recency of task performance. All specialties were aggregated for these analyses to obtain sufficient sample sizes for each aptitude by experience combination. The proficiency measures for both the HOPTs and JKTs were limited to the core infantry tasks and items that were administered to all MOSs.

Time in Service

Figure I shows the plot of average hands-on proficiency scores for both high- and low-aptitude personnel (GT < 100 versus GT => 100) at yearly intervals. Error bars are noted for each mean; areas in which the error bars do not overlap indicate significant proficiency differences between groups. Figure I illustrates several important findings:

- For both aptitude groups, hands-on proficiency showed steady advancement through the first three years of service, after which time such progress lessened.

- The high-aptitude group always outperformed the low-aptitude group at each yearly interval.

- Marines of high aptitude who were in their first year of service were able to perform at the same level as low-aptitude Marines with more than three years of service.

Although low-aptitude Marines benefited from training and on-the-job experience and improved their proficiency over time, high-aptitude Marines also bettered their proficiency. The initial proficiency level of high-aptitude Marines was substantially higher than that of the low-aptitude Marines. This initial proficiency gap was so large that three or more years of experience were required for a low-aptitude Marine to accomplish the same proficiency level of a high-aptitude, first-year Marine.

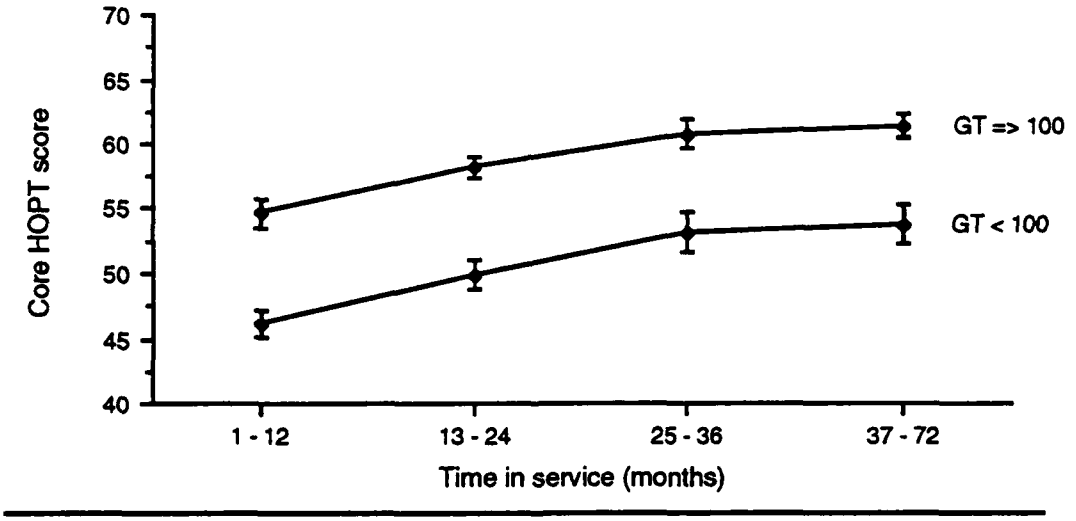

Figure I. Mean core hands-on performance by time in service and aptitude level

Rank

Regardless of the aptitude group, higher ranks were able to perform significantly better than their subordinates (see figure II). As with time in service, the higher aptitude group was able to achieve significantly higher hands-on proficiency levels than the low-aptitude group for each rank category. Some interesting tradeoffs between aptitude and rank can be noted.

- Comparable hands-on proficiency levels were achieved by high-aptitude lance corporals (LCpls—pay grade E3) and low-aptitude corporals and sergeants (Cpl/Sgts—pay grades E4 and E5).

- Privates and privates first class (Pvt/PFCs—pay grades E1 and E2) of high aptitude performed only slightly less than Cpl/Sgts of low aptitude.

- Notable proficiency differences occurred between high-aptitude Pvt/PFCs and low-aptitude LCpls; the Pvt/PFCs performed significantly better.

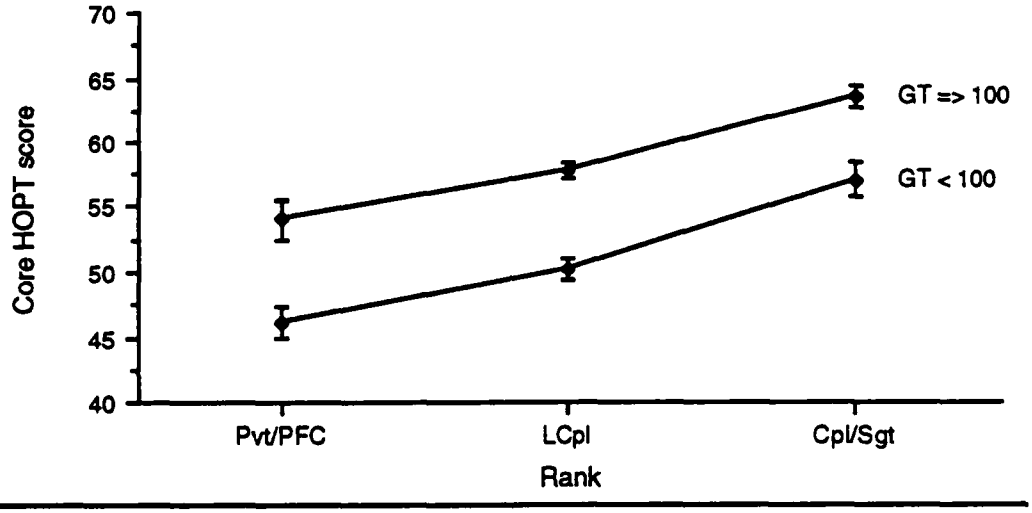

Figure II. Mean core hands-on performance by rank and aptitude level

Despite increasingly more difficult job requirements for higher ranks, high-aptitude personnel at lower ranks are better able to assimilate complex concepts or to model their superiors' performance so as to significantly exceed the proficiency of higher ranking but lower aptitude Marines. Although the hands-on tests did not directly measure leadership or the ability to work in small units, stark differences were revealed in overall proficiency to perform critical infantry tasks relative to aptitude and rank. Such comparisons of proficiency by rank are useful in developing the case for high-quality recruits because it is both an experience and performance metric readily understood by and operationally meaningful to the Marine Corps.

Recency of Task Performance

Figure III shows that recency of task performance had a positive effect on hands-on proficiency for both aptitude groups. Those persons with more recent task experience (within the past month) performed significantly better than those who had limited practice (greater than one month earlier) or who had instruction only and no experience. Figure III also reveals that the hands-on proficiency differences between high- and low-aptitude groups were not significant for the "instruction only" condition. Thus, for training settings in which individuals were merely exposed to the material and had no opportunities to perform the task, differences in hands-on proficiency were not significant for high- versus low-aptitude groups. However, given "limited" opportunities to perform the infantry materials, the high-aptitude group showed significant gains in hands-on proficiency. The low-aptitude group showed only slight improvement but not to a significant degree.

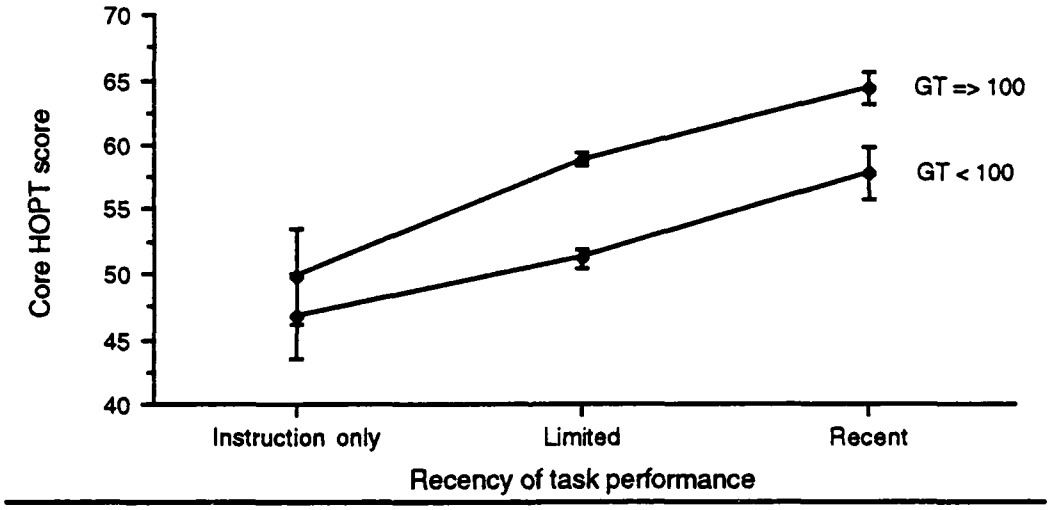

Figure III. Mean core hands-on performance by recency of task performance and aptitude level

Several general implications concerning infantry training can be drawn from these findings. First, a certain standard of performance can be attained by the high-aptitude group with substantially less training or on-the-job experience than would be required by the low-aptitude group. Figure III shows that high-aptitude persons need only limited experience in task performance to perform as well as low-aptitude persons with much more recent task exposure. Second, significantly higher proficiency levels can be achieved by a high-aptitude group for a fixed level of training. Figure III shows that, given only limited performance opportunities, the

high-aptitude group was about 16 percent more proficient than the low-aptitude group. The tradeoffs between aptitude and recency of task performance demonstrate that recency of performance does not compensate for low initial aptitude.

CONCLUSIONS

The Marine Corps selected the infantry occupational field for the initial JPM testing because it is the largest and often the most controversial for establishing aptitude requirements. More than 2,300 infantrymen in five MOSs were tested. The findings for each MOS were particularly noteworthy and provide the Marine Corps with a solid case for definitively illustrating the effect of recruit quality on infantry performance. ASVAB was found to be highly valid as a predictor of several different infantry performance measures: hands-on performance of critical infantry tasks, a written job knowledge test, and grades in training courses. Such results strongly dispute the common belief that anyone, regardless of their aptitude level, can function equally well as an infantryman. Aptitude was shown to be highly relevant to the successful performance of infantry job requirements.

The strong relationship between aptitude and performance also extended into the second term of service. Individuals with higher aptitudes were found to be the better second-term performers. This outcome has significant implications for the types of individuals the Marine Corps should recruit to be able to staff its leadership positions with qualified personnel in future years. In addition, these results indicate that the Marine Corps would have difficulty in quickly recovering from the detrimental effect of low recruit quality. Declines in quality would have substantial long-term effects by reducing not only the level of first-term performance but also eventually affecting the caliber of enlisted leadership.

Experience in the Marine Corps, as measured by time in service, rank, and recency of task performance, had a significant and positive influence on job proficiency. Despite such developmental effects, experience was not able to compensate for low aptitude in the prediction of proficiency. The initial gap in hands-on proficiency between high- and low-aptitude groups was so great that more than three years of service or the rank of at least corporal were required to offset the deficit. The recency of task performance ratings also showed that significantly more training or practice was required of lower aptitude personnel to achieve the same proficiency outcomes as those of higher aptitude personnel with less training or practice.

Although these validity results provide convincing evidence for the impact of aptitude on performance, the Marine Corps must address a secondary issue concerning what performance levels it requires from its infantrymen. Significant validities imply that "more aptitude is better," but it does not confront the issue of "how much performance is enough." This determination of required performance levels is an important ingredient in the Marine Corps' overall justification of its requirements for recruit quality. Establishing the aptitude/performance linkage is an initial and necessary, but not sufficient, condition in this process.

The Marine Corps JPM project succeeded in developing reliable, objective, and representative measures of infantry performance. The hands-on performance tests did not necessarily measure an individual's ability to lead, to operate in stressful or hostile environments, or to work effectively with others in small units. Also, the performance tests were not collected under combat-like conditions (the ultimate performance criterion); however, combat would only degrade the performance levels observed in the JPM project. Individuals currently not able to accomplish the requisite infantry tasks in a nonthreatening environment probably would not be able to perform them under the chaotic conditions of war. The JPM project has empirically demonstrated a strong and persistent relationship between individuals' aptitudes and their future performance in the infantry occupational field and, ultimately, their ability to function more effectively in combat by being able to perform critical infantry skills.

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

TABLES

INTRODUCTION

The joint-service Job Performance Measurement (JPM) project was initiated in 1981 to link military aptitude requirements to objective measures of job performance. Because of problems during the 1976-1980 period with the Armed Services Vocational Aptitude Battery (ASVAB), the instrument used for selecting and classifying military recruits, the historical criterion of training grades was questioned. At that time, because of a misnorming error, significant numbers of unqualified personnel entered the services because ASVAB scores significantly overestimated individuals' true aptitudes. The impact on military effectiveness of such low-aptitude and unqualified personnel could not be established empirically. The JPM project was charged with accurately measuring the job performance of military personnel and directly relating this performance to the ASVAB.

The Marine Corps has focused on the infantry occupational field as its contribution to this joint-service project. This research memorandum examines several aspects of the Marine Corps project that have implications related to manpower, recruiting, and training:

- Establishing the relationship between ASVAB and infantry job performance
- Examining differences in validities for each of the Marine Corps aptitude composites
- Investigating the interaction between aptitude and experience in predicting infantry performance.

The Marine Corps JPM project has focused on the objective assessment of infantry job performance. Early in the project's history, the joint-service JPM Working Group established hands-on performance tests (HOPTs) as the benchmark measure of job performance [1]. HOPTs have high fidelity to actual job performance and require the examinee to perform job tasks under realistic but standardized testing conditions. In addition, job knowledge tests (JKTs) were developed for the Marine Corps project. The JKTs were designed to be parallel in test content to the HOPTs and included certain knowledge items related to job performance that could not be tested in the hands-on mode. Marine Corps proficiency marks (PRO marks), supervisors' assessment of Marines' infantry performance, were also obtained. Training grades, the historical validation criterion, were also collected for each Marine to examine their relationship with ASVAB. The development and administration

of both the HOPTs and JKTs are briefly discussed in the next section, along with the procedures used to collect PRO marks and training grades.

The term "job proficiency" is used in this research memorandum to reflect an individual's ability to accomplish selected job tasks in a standardized testing situation, i.e., performance on the hands-on tests or job knowledge tests. Conversely, the term "job performance" is used to mean what individuals generally do on their jobs and the assessment of individual ability in a generic sense.

TEST DEVELOPMENT AND ADMINISTRATION

Hands-On Performance Test

The initial requirement in developing a hands-on performance test is to specify completely the tasks that constitute a job. The Individual Training Standards (ITS) developed by the Training Department, Headquarters, Marine Corps, were the primary source of detailed information that defined the job tasks of the infantry occupational field [2]. The ITS tasks are organized into relatively homogeneous duty areas. Table 1 shows the duty areas of the basic infantry military occupational specialty (MOS) and examples of tasks from the ITS. In addition to these core infantry tasks required of all infantry specialties, the ITS specify job requirements that are unique to each of the other infantry specialties.

Table 1. *Duty areas and examples of hands-on tasks for basic infantry MOS*

Duty area	Example of hands-on task
Tactical measures	Call for/adjust indirect fire
Security and intelligence	Process prisoners and equipment
M203 grenade launcher	Prepare for firing
Hand grenades	Engage target with dummy grenades
Mines	Install Claymore mine
Communications	Assemble and operate PRC-77 radio
Land navigation	Determine location by resection
First aid	Treat sucking chest wound
Night vision device	Operations inspection
Squad automatic weapon	Fieldstrip and assemble
Light antitank weapon (LAW)	Restore expanded LAW
Nuclear, biological, chemical defense	Don protective clothing

Because of limited resources, personnel, and time, not all tasks defined by the ITS could be tested; therefore, an objective sampling procedure had to be developed [3]. The sampling procedure incorporated the underlying behavioral similarities needed to perform all infantry tasks. Each task was weighted by its number of behavioral elements, and the tasks were then randomly sampled within each duty area. The intent was to test as many behaviors as possible within a duty area without being redundant in the testing of any specific behaviors.

This task sampling process was employed so that scores resulting from the hands-on test would be representative of a Marine's ability to perform all infantry tasks based on his performance of only the sampled subset. The score scale for the hands-on test was a percent-correct metric so that meaningful interpretations could be made of differences in individual proficiency. For example, a hands-on score of 90 implies that a Marine could successfully perform 90 percent of all infantry job requirements. Also, the hands-on tests were developed so that scores would have ratio scale properties: a Marine scoring at the 90-percent level is 1.5 times more proficient than a Marine with a hands-on score of 60.

Marine Corps job experts were involved extensively throughout the task specification and sampling process. The job experts ensured that the selected test content represented infantry specialties as a whole and was consistent with what is required on the job and in combat. The tasks were then analyzed to identify discrete and observable steps associated with task performance. From these task analyses, hands-on tests were created with performance steps that could be scored objectively in a pass/fail format. To minimize the ambiguity and subjectivity associated with scoring hands-on tests, distinct performance behaviors were identified and definitive scoring criteria established. These hands-on measures were tried out on current job incumbents to refine test administration and scoring procedures and to ensure that testing materials were representative of actual job performance.

The hands-on test for each infantry MOS included both a core set of infantry tasks and unique tasks specifically required of that specialty. Some of the unique MOS requirements included: riflemen firing the M16A2 rifle on a pop-up quick-response range, machinegunners preparing range cards for the Mk 19 machinegun, mortarmen laying the 81mm mortar, assaultmen firing the launch effects trainer (a simulator for the Dragon antitank missile), and infantry unit leaders directing close-air support delivered by fixed-wing aircraft. The hands-on test for each MOS required about eight hours to complete.

The test administrator is the most critical component of hands-on testing. Unlike paper-and-pencil tests in which reliable and objective scoring keys are readily applied, hands-on testing involves a judgment of whether an individual performed a particular action. To minimize the subjectivity involved, the test administrators were trained for two weeks in the performance of all tasks and in procedures for the standardized administration of tests. Retired Marine Corps staff noncommissioned officers and officers were hired as test administrators because of their experience in the infantry field, knowledge of the Marine Corps, and ability to work well with young Marines. To ensure comparability of hands-on scoring standards across testing locations, detailed training materials were prepared, and the same training team conducted the instructional sessions at both testing locations.

Extensive quality-control procedures were implemented to ensure that the test administrators maintained the scoring standards to which they were originally trained. These control measures included multiple-administrator scoring of examinee performance to determine administrator agreement and consistency, daily computerized entry of performance data to check for administrator leniency or drift over time, and test administrator rotation across testing stations to minimize systematic error.

Job Knowledge Test

A job knowledge test was developed for each infantry MOS to be parallel in test content to the hands-on tests. Development of the JKT items began with a review of the steps of the hands-on tests that were crucial to the performance of the task. Items were then written to capture these critical aspects of task performance. To make the written items as performance-based as possible, extensive use was made of graphic materials and illustrations. For tasks that did not involve weapons (e.g., tactical measures, land navigation), combat scenarios were composed and Marines were asked what actions were appropriate for the given situation.

The JKTs were pilot-tested on a sample of infantry riflemen. The pilot test included more items than were needed so that items of poor quality could be deleted. Between 190 and 200 items were chosen for the final JKT for each MOS. Two hours were allowed to complete the test.

Proficiency Marks

Proficiency marks (PRO marks) are the Marine Corps' operational-supervisor ratings reflecting how well a Marine performs his job. PRO marks are received about every six months, or earlier if an individual is transferred to another unit. These ratings enter into a composite score used for promotion decisions and occupational field transfers. PRO marks range from 0.0 to 5.0, but the actual score distribution is much smaller: about 90 percent of all Marines receive marks between 4.2 and 4.8.

The PRO mark score used for these analyses was the mean of all available proficiency marks for an individual. Over 90 percent of Marines in the JPM sample had received at least three PRO marks in their career. (This does not include marks given at the conclusion of the School of Infantry.) On average, each person had received more than five supervisor ratings. The scale for the PRO marks used throughout this research memorandum has been multiplied by 10 to eliminate decimals.

Training Grades

Grades in training have been the historical criterion against which the ASVAB has been validated. Final grade-point average (GPA) in the School of Infantry was collected for each Marine tested, if available. Because all infantry MOSs complete the same initial training, examinees were not analyzed separately by MOS.

Testing Process

Headquarters, Marine Corps, randomly sampled the Marines to be tested from the available units for each MOS. To ensure that the sample was representative of the population of infantrymen, the sampling was stratified by rank, educational level, and length of service. Based on these selection criteria, the desired sample sizes were:

0311	Rifleman	1,000
0331	Machinegunner	300
0341	Mortarman	300
0351	Assaultman	300
0369	Infantry unit leader	400

Alternate Marines with similar characteristics were also sampled to replace personnel who might be unavailable for testing. Of the 1,000 riflemen, about 200 individuals were retested with the alternate form of the proficiency tests to determine the reliability of the testing procedures.

Each Marine was tested for two days. The hands-on and written tests were organized into testing blocks and staggered so that the Marines were not required to do all paper-and-pencil exercises contiguously. Other performance information such as PRO marks and grade-point average from the School of Infantry were also collected for each Marine.

Despite significant precautions and quality-control procedures, some individual cases of incomplete or inaccurate data still occurred. Reference [4] documents the decision rules and procedures used to impute missing data and the methods used to identify inconsistent response patterns.

CRITERION RELIABILITY

If a performance criterion is unreliable and contains measurement error, estimates of validity are likewise affected. Theoretically, a performance test cannot be more highly related to another measure than it is to its own true score (a test score measured without error). Criterion reliability must be a primary consideration before validity is examined. Several reliability estimates were computed:

- Test-retest reliability: comparison of test performance on alternate test versions. About a 20-percent sample of the infantry riflemen were retested after seven to ten days with an alternate test form.

- Alpha coefficient: a measure of the internal consistency of test items (or tasks) that reflects the degree to which items measure the same concept.

- Scorer agreement: the percent agreement between two test administrators as they score the step-level performance of one examinee.

These reliability estimates were computed, as appropriate, for the HOPTs, JKTs, and PRO marks. All reliability estimates were corrected for range restriction. Sample reliability estimates are reported in appendix A.

Despite the potential for many sources of error, the hands-on tests were found to be very reliable (see table 2). Test-retest reliability was .77. Alpha coefficients were consistently high for the HOPTs of all MOSs. Additional analyses designed to examine other factors influencing the variance of hands-on tests showed that experienced and well-trained test administrators can dependably score proficiency across time, different persons, and varying test content [5].

Table 2. Reliability estimates[a] for hands-on performance tests and job knowledge tests

Reliability measure	MOS				
	0311	0331	0341	0351	0369
HOPT					
Test-retest	.77				
Alpha coefficient	.90	.91	.89	.85	.89
Scorer agreement	.92	.93	.90	.91	.92
JKT					
Test-retest	.85				
Alpha coefficient	.94	.93	.92	.91	.93

a. Reliability estimates are corrected for restriction in range.

Figure 1 presents the scatterplot of the hands-on retest versus the initial test. The diagonal line indicates the line of no change in test performance across the two testing sessions. Points above this line represent persons who improved their performance relative to their initial test; points below the line represent persons whose retest performance was lower than their initial performance. A significant retest gain occurred in hands-on proficiency of over .8 standard deviation (SD). Such gains noted seven to ten days later may reflect the positive impact of practice on the performance of infantry tasks or simply a better understanding of the hands-on testing procedures. Further analysis of these retest improvements showed that the gains were not related to aptitude; both high- and low-aptitude personnel made equivalent advances in hands-on proficiency.

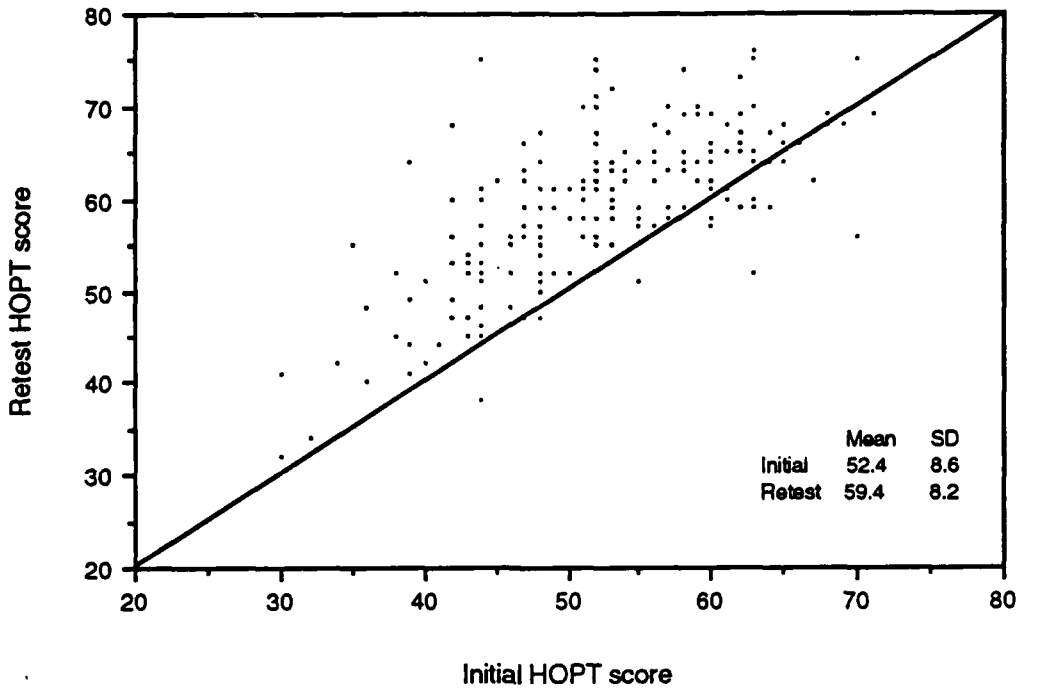

Figure 1. Retest reliability for hands-on performance test

Table 2 shows that the written job knowledge test was slightly more reliable than the hands-on measures. Figure 2 depicts a test-retest reliability of .85 for the JKT with no average retest gain. Practice appears to have had less effect on job knowledge than on hands-on proficiency. Job knowledge may be stable and not influenced by repeating performance after a seven- to ten-day interval. The JKTs were found to be rather difficult tests; on average, infantrymen answered about 45 percent of the written items correctly. The alpha coefficients ranged from .91 to .94 for the five MOSs.

The reliability of PRO marks was computed based on an analysis of variance design. Estimates were computed for the three, four, and five most recent ratings that an individual had received. The reliabilities ranged from .67 to .71, implying that the marks are relatively stable and consistent for a variety of conditions (e.g., across time, for different raters).

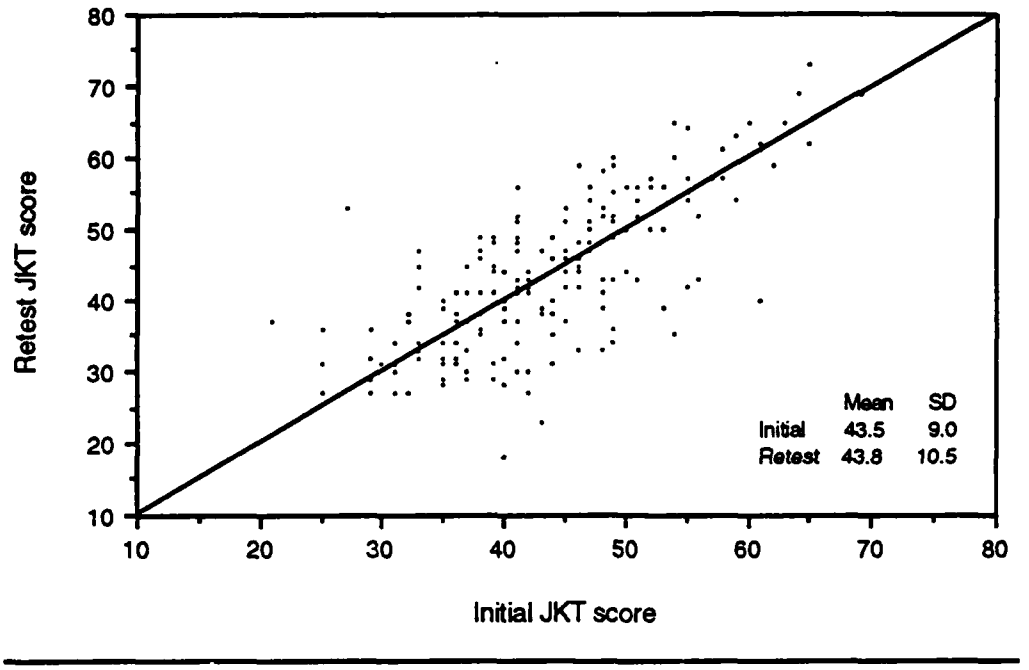

Figure 2. Retest reliability for job knowledge test

ASVAB VALIDITY

The ASVAB is composed of ten subtests, which the Marine Corps combines into four aptitude composites: general technical (GT), mechanical maintenance (MM), electronics repair (EL), and clerical/administrative (CL). These composites are used to classify recruits into occupational specialties for which they will have the greatest likelihood of success. The GT aptitude composite is used to determine eligibility for MOSs of the infantry occupational field. A validity coefficient expresses the relationship between an ASVAB aptitude composite and a performance measure. If the aptitude composite is related to performance, the use of ASVAB is then warranted as a selection and classification test.

Table 3 reports the validities of the GT composite against HOPTs, JKTs, and PRO marks. The validities have been corrected for range restriction based on a multivariate procedure that considers the effects of selection based on all ASVAB subtests [6]. Sample validities are reported in appendix B along with descriptive statistics for each variable.

Table 3. Validity[a] of GT composite against infantry *performance measures*

MOS	Sample size	Performance measure		
		HOPT	JKT	PRO
Rifleman	940	.62	.78	.32
Machinegunner	271	.70	.74	.22
Mortarman	253	.48	.65	.05
Assaultman	277	.51	.68	.11
Infantry unit leader	387	.60	.73	—[b]

a. Validities were corrected for range restriction.
b. PRO marks are not given to infantry unit leaders.

The corrected validity coefficients for the GT composite and HOPTs ranged from .48 to as high as .70. These results indicate that there is a strong relationship between how well an individual performed on the ASVAB and his later performance as an infantryman. A second significant outcome of the results in table 3 is the magnitude of the validity coefficient for second-term infantry unit leaders, indicating the robustness of aptitude in predicting second-term proficiency. Unit leaders with higher aptitudes were better performers than those with lower aptitudes. Both of these findings strongly support the continued use of ASVAB in the selection and classification of infantry recruits.

Validities of the GT composite against JKT scores were somewhat higher than those for HOPTs (.65 to .78), but substantially lower against PRO marks (.05 to .32). The results for the JKTs are comparable to previous validation research using training grades as the criteria [7, 8]. The lower results noted for the PRO marks across MOSs were also noted in other studies [9] and may be due to the subjectivity of supervisor ratings or lack of PRO mark variability. Again, the results for JKTs provide strong empirical support for the validity of the ASVAB.

The Marine Corps conducts infantry training at two primary locations. Previous analysis has shown different relationships exist between aptitude and grades depending on training location [8]. Table 4 presents the validities for each

aptitude composite separately by training location, which confirm that earlier research. Base A had consistently high validities for each aptitude composite against infantry GPA. Conversely, the validities for Base B were lower by about 17 points for each composite.

Table 4. Corrected validity for School of Infantry training grades[a] by training location

Aptitude composite	Base A	Base B
GT	.61	.43
MM	.61	.42
EL	.60	.43
CL	.53	.38
AFQT	.58	.41

a. MOSs were aggregated because individuals complete the same initial infantry training.

The distribution of GPA at various aptitude levels showed much more systematic variance in GPA at Base A, whereas significant numbers of persons achieved the maximum possible GPA at Base B (even individuals with low aptitude). These differences in GPA distributions possibly reflect differences in training philosophies between the two locations. Base B may follow the Instructional System Development (ISD) procedures for training more closely, with individuals progressing through instructional blocks based on pass/fail criteria. Given enough time, potentially everyone passes the course. Such grading systems that do not include information about how long individuals take to complete the course are not conducive to validation research. Nevertheless, the validities for Base B were still moderately high. The results for Base A provided strong evidence for the validity of ASVAB in predicting training outcomes. Likewise, these findings confirm the Marine Corps historical use of training criteria for validation purposes.

Scatterplots of Hands-On Performance Tests

Figure 3 shows the scatterplot and regression of the HOPTs on the GT aptitude composite for the infantry rifleman specialty. Consistent scatter occurred about the regression line without any apparent patterns or significant outliers.

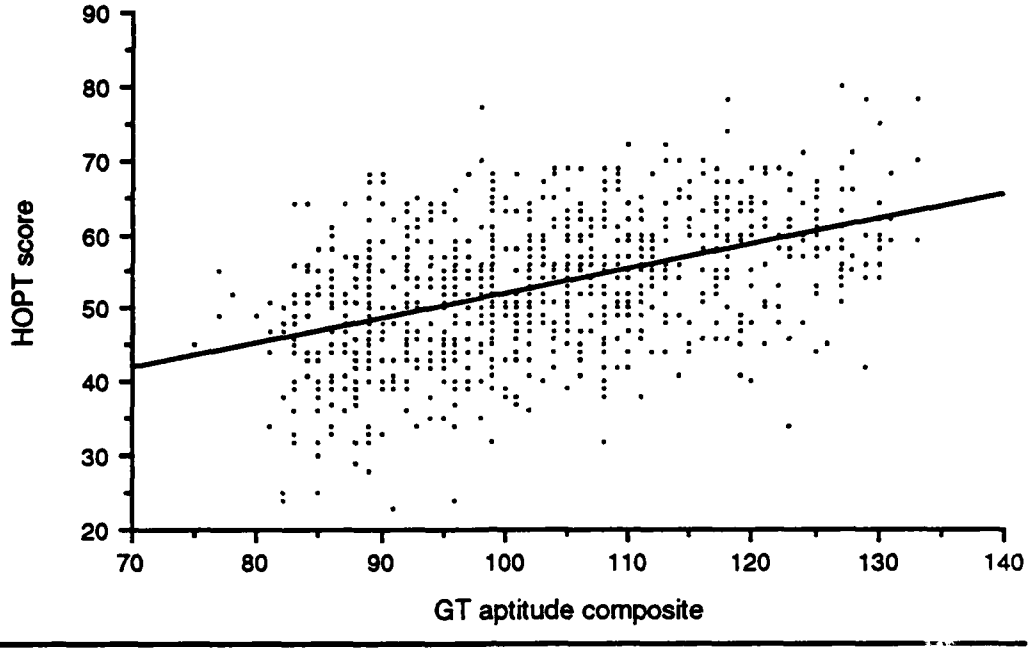

Figure 3. Scatterplot of hands-on performance by GT aptitude composite for infantry riflemen

The assumptions for correcting validity coefficients due to range restriction require that the regression be linear throughout the predictor range and that the variance of errors be constant for all values of the predictor (homoscedasticity). These assumptions are best tested by examining a scatterplot of the standardized residuals versus the GT aptitude composite (see figure 4). The expectation is that less than 1 percent of the observations (about 9 points) should, by chance, fall outside the range of plus or minus 3 standard residuals. The plot shows only a few such data points outside this range. Also, no trends were found in the residuals to imply the presence of a significant quadratic or cubic component to the regression (i.e., the regression is linear); therefore, the corrections made to validity coefficients due to range restriction are warranted and appropriate.

Performance Differences Associated With Validity

Although the validity coefficients indicate a strong relationship between aptitude and performance on the hands-on and job knowledge tests, the coefficients do not show how much personnel of varying aptitude differ in their performance. High validity coefficients imply that high-aptitude personnel perform better, but how much better? Figure 5 shows the hands-on proficiency means of infantry riflemen at different score intervals of the GT composite. Each of the GT intervals contains more than 175 examinees, so sufficient sample sizes were available to compute stable means.

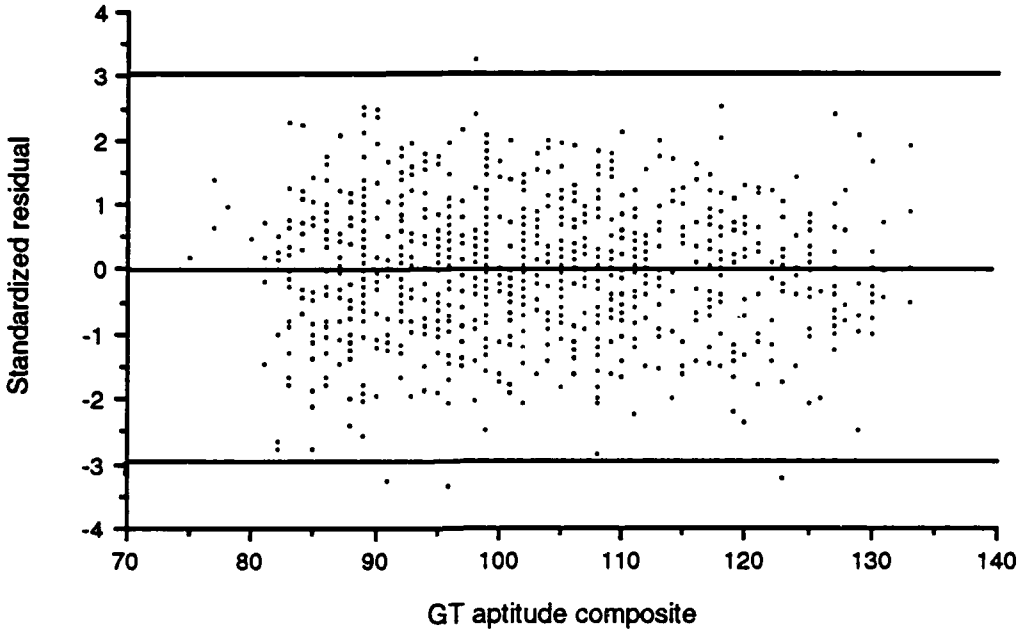

Figure 4. Plot of standardized residuals vs. GT aptitude composite for infantry riflemen

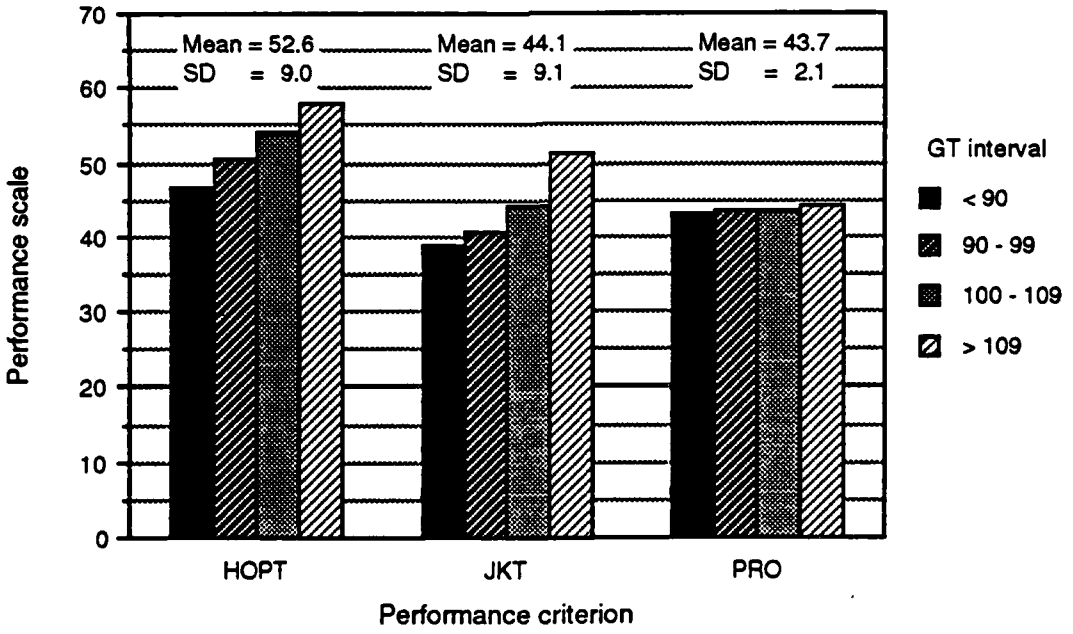

Figure 5. Performance means of infantry riflemen for GT intervals

Individuals with GT scores greater than 109 performed about 25 percent better on the HOPTs than individuals with aptitudes less than 90.[1] Increases in hands-on proficiency showed steady improvements of about 4 points for every 10-point increment in GT. Based on the median of the GT intervals, changes of 10 to 12 percent in GT yielded a 7- to 8-percent improvement in mean hands-on proficiency (see table 5).

Table 5. Percent of change in GT intervals vs. percent of change in mean infantry rifleman performance measures

GT interval (median)	Sample size	Percent of change in mean performance to achieve next higher GT level			
		GT	HOPT	JKT	PRO
< 90 (85)	178				
		12	8	4	1
90 - 99 (95)	274				
		11	7	9	0
100 - 109 (105)	225				
		10	7	16	1
> 109 (115)	263				
Highest group versus lowest group		35	25	31	2

1. Comparison of performance differences between high- and low-aptitude groups are based on the ratio scale properties of the hands-on and job knowledge tests. The percentage differences are computed using the low-aptitude group as the base (i.e., percentage difference = (high-low)/low).

A similar trend of proficiency improvements is also noted in figure 5 for the JKTs. Although the overall JKT proficiency difference between the lowest and highest aptitude groups was over 30 percent, the largest gain was noted at aptitude levels greater than 109. Table 5 shows that improving aptitude by 10 points at the lower end of the GT scale returned only slight JKT gains of about 4 percent versus about a 16-percent improvement when the GT composite score was increased from 105 to 115. Thus, while aptitude was significantly related to JKT proficiency, the return on an investment in quality would be even greater at higher aptitude levels.

The rather flat profile of means for PRO marks across all GT intervals illustrates the lack of relationship between aptitude and supervisor ratings that was discussed earlier. Persons with high GT scores received PRO marks that were only 2 percent higher than persons with the low GT scores.

DIFFERENTIAL VALIDITY OF APTITUDE COMPOSITES

The Marine Corps uses four aptitude composites to classify all recruits into occupational areas for which they have the greatest likelihood of success.[1] The fundamental assumption behind using different composite scores for classification purposes is that the composites can predict performance differentially across a variety of Marine Corps jobs. That is, persons with high GT scores will become good infantrymen but not necessarily be good mechanics; persons with high CL scores will become good clerks but not necessarily be good electronics repairmen. Only if such differential validity exists across jobs is the continued use of distinct composite scores justified.

Figure 6 plots the corrected validities for the hands-on proficiency of each infantry specialty against the aptitude composites. The corrected validities from which these plots were constructed and the sample validities for each composite are reported in appendix B.

Although the GT composite is used for classification into infantry specialties, the MM composite was a better predictor of infantry hands-on proficiency, particularly for the rifleman and mortarman specialties. The validity of the EL composite was comparable to MM validity for the machinegunners and assaultmen. In all cases, the validity for the CL composite was substantially lower than the validity for

1. The Armed Forces Qualification Test (AFQT) is another important composite, but one that is not explicitly used by the Marine Corps for classification purposes. Rather, the AFQT is a measure of general trainability and is used to establish recruiting goals.

any other composite in predicting hands-on proficiency. The CL composite includes one of the timed subtests of the ASVAB. These results indicated that perceptual speed is not highly related to the successful performance of infantry tasks. The AFQT also was less related to the HOPTs than were GT, MM, or EL.

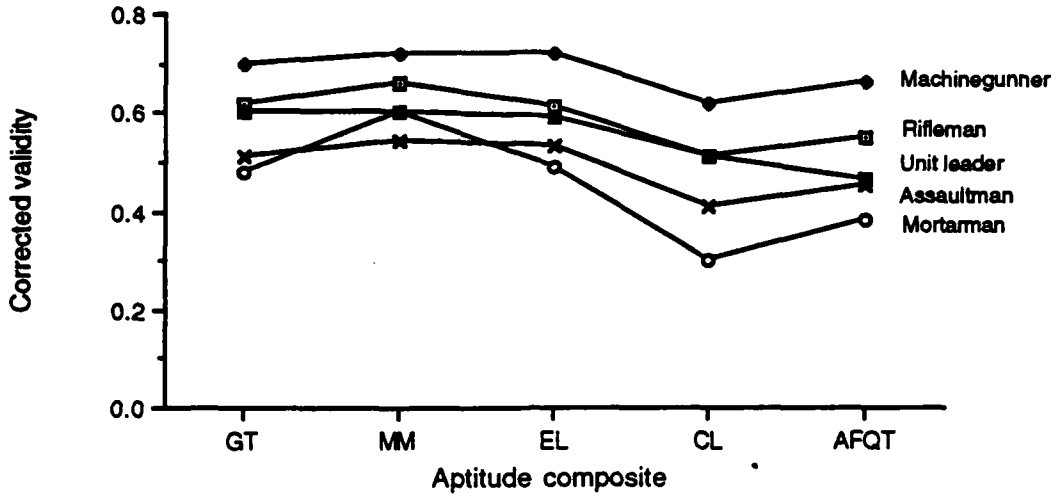

Figure 6. Differential validity of aptitude composites vs. hands-on performance test

Higher overall validities were found for each of the composites when JKTs were the proficiency measures. This outcome was due primarily to the common testing medium (i.e., both the ASVAB and JKTs are written tests—persons who do well on written selection tests also tend to do well on written performance tests). Figure 7 shows less difference in validities across the aptitude composites versus JKTs than versus the HOPTs. Contrasted to the findings for the hands-on test, the MM composite against JKTs was one of the worst composites (except for mortarmen), and CL predicted infantry proficiency almost as well as the other composites. (Again, mortarmen were an anomaly.)

Validities against PRO marks were computed and plotted in figure 8. These validities were a magnitude lower than those for the other criteria and showed virtually no change across the aptitude composites.

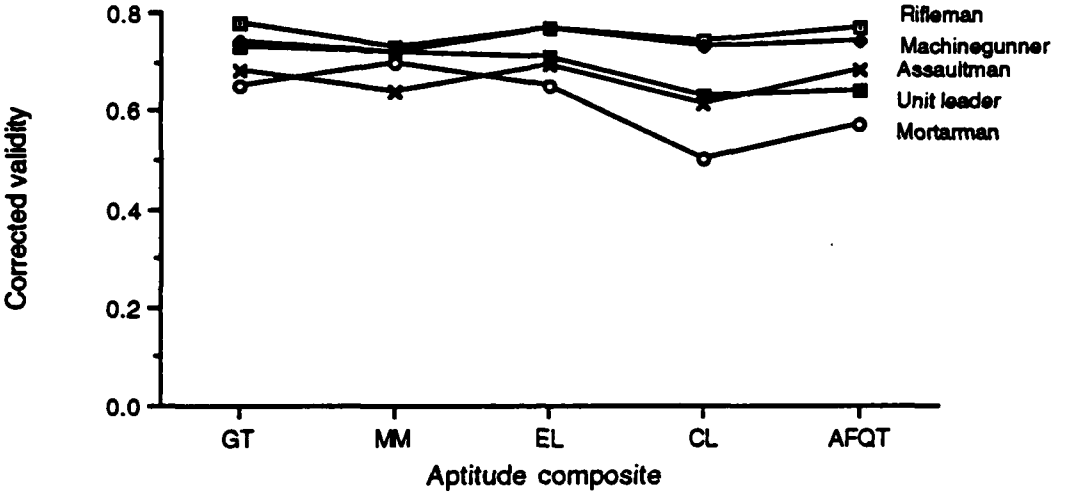

Figure 7. Differential validity of aptitude composites vs. job knowledge test

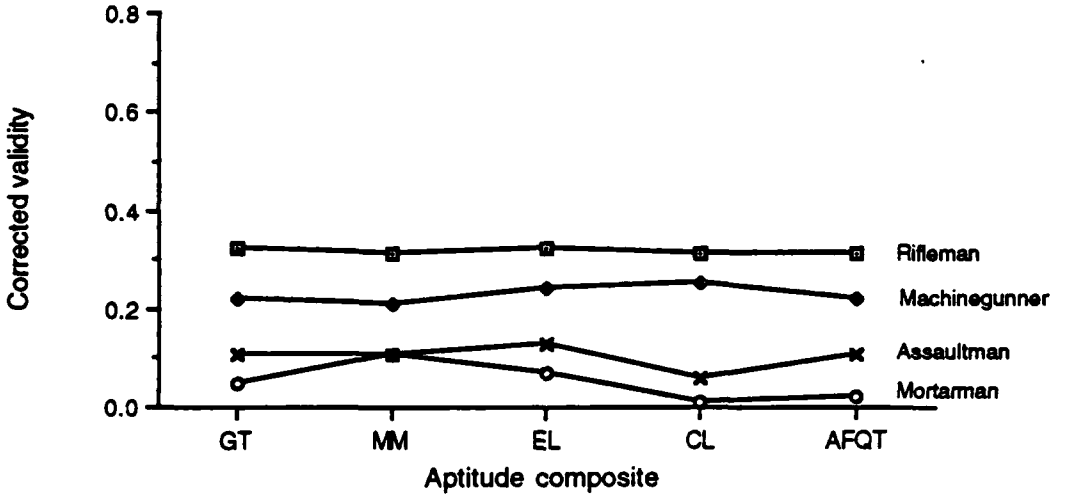

Figure 8. Differential validity of aptitude composites vs. proficiency marks

Given that the ASVAB has been validated historically against readily available criteria such as training grades (based primarily on written tests similar to the JKTs) and supervisor ratings, it is not surprising that researchers have found the ASVAB to have limited differential validity [10]. The findings for PRO marks (and JKTs to a lesser extent) show a lack of variation in validity across the composites. However, hands-on performance tests have been established by the joint-service JPM Working Group as the benchmark performance criterion against which the ASVAB should be validated. Findings for the HOPTs demonstrate that the ASVAB does have differential validity against a high fidelity performance measure.

Although these results provide empirical support for differential validity by aptitude composite, they also indicate that the Marine Corps could benefit by using a composite more like MM than GT to classify recruits into infantry specialties. Before implementing such findings, however, additional investigation is required to determine if unacceptable implications are associated with such a change (e.g., competition with other MOSs for individuals with high MM scores).

INTERACTION OF APTITUDE AND JOB EXPERIENCE

The relationship between hands-on performance tests and job experience is important for establishing the measurement properties of hands-on performance tests. The expectation is that hands-on performance should increase with experience as job incumbents acquire higher levels of proficiency through on-the-job training and more advanced instruction. Also, if aptitude is a valid predictor of proficiency, personnel with high aptitude should outperform their counterparts with low aptitude. The interesting question involves the interaction of both of these variables and the prediction of proficiency: Do high- and low-aptitude individuals differ in their performance across all levels of experience or can experience compensate for lower aptitude? This latter issue potentially has significant implications for Marine Corps recruiting—quality goals could possibly be reduced if lower aptitude recruits could be trained to perform as well as higher aptitude recruits.

The interaction of aptitude and experience was examined for three measures of experience:

- Time in service
- Rank, which incorporates Marine Corps recognition of performance through promotion and time in service
- Recency of task performance, a self-report measure of experience reflecting recency of opportunity to perform the specific tasks that were tested.

Each of these definitions of experience was examined for an interaction with aptitude in the prediction of hands-on and job knowledge test proficiency.

The five infantry specialties were aggregated for these analyses to obtain reasonable sample sizes for each aptitude by experience combination. The proficiency measures for both HOPTs and JKTs were limited to the core infantry tasks

and items that were administered to all MOSs. This common infantry content represented over 75 percent of all testing material. Descriptive statistics of the core infantry portions of the HOPTs and JKTs are reported in appendix C for each aptitude/experience combination.

Time in Service

Figure 9 plots the mean hands-on proficiency for both high- and low-aptitude personnel (GT < 100 versus GT => 100) at yearly time-in-service intervals. A GT score of 100 represents the aptitude score of an average recruit from the 1980 youth population; the rifleman specialty had a mean GT score of 102. Error bars are noted for each mean (plus or minus 2 standard errors of the mean). Areas in which the error bars do not overlap indicate significant proficiency differences between groups.

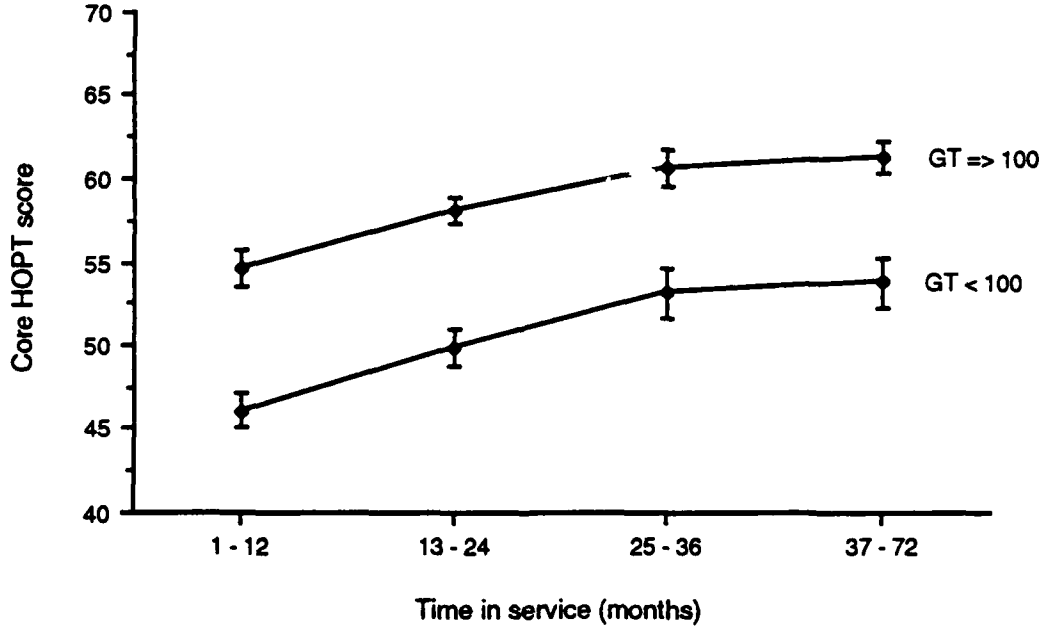

Figure 9. Mean core hands-on performance test by time in service and aptitude level

A number of important findings are apparent from figure 9. First, hands-on proficiency showed steady advancement through the first three years of service, after which time such progress lessened. The hands-on proficiency of Marines in their third year of service was not significantly different from the proficiency of

more senior Marines. This finding was consistent for both aptitude groups. Second, the high-aptitude group always outperformed the low-aptitude group at each yearly interval. Finally, Marines of high aptitude who were in their first year of service were able to perform at the same level as low-aptitude Marines with more than three years of service. Thus, although low-aptitude Marines benefited from training and on-the-job experience and improved their proficiency over time, high-aptitude Marines also bettered their proficiency; however, high-aptitude Marines' initial proficiency levels were substantially higher than those of low-aptitude Marines. This initial proficiency gap was so large that three or more years of experience were required for a low-aptitude Marine to accomplish the same proficiency level as that of a high-aptitude, first-year Marine.

The identical analysis was conducted using the core JKT as the criterion, and essentially the same outcomes were found (see figure 10). The leveling off did not occur for JKT proficiency that was noted for HOPT proficiency; JKT proficiency steadily progressed over the years. Again, the first-year proficiency difference between the high- and low-aptitude groups was of such magnitude that more than three years of service were required to compensate for this deficit.

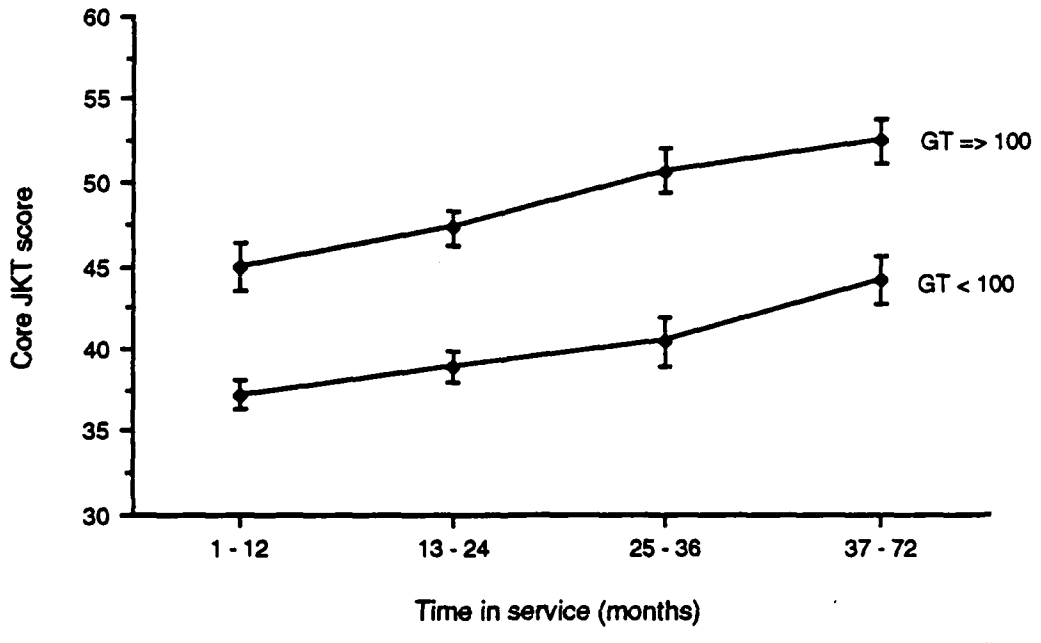

Figure 10. Mean core job knowledge test by time in service and aptitude level

Rank

Examinees were grouped into one of three categories according to rank: private and private first class (Pvt/PFC—pay grades E1 and E2), lance corporal (LCpl—pay grade E3), and corporal and sergeant (Cpl/Sgt—pay grades E4 and E5). The mean HOPT proficiency for each rank category was plotted in figure 11.

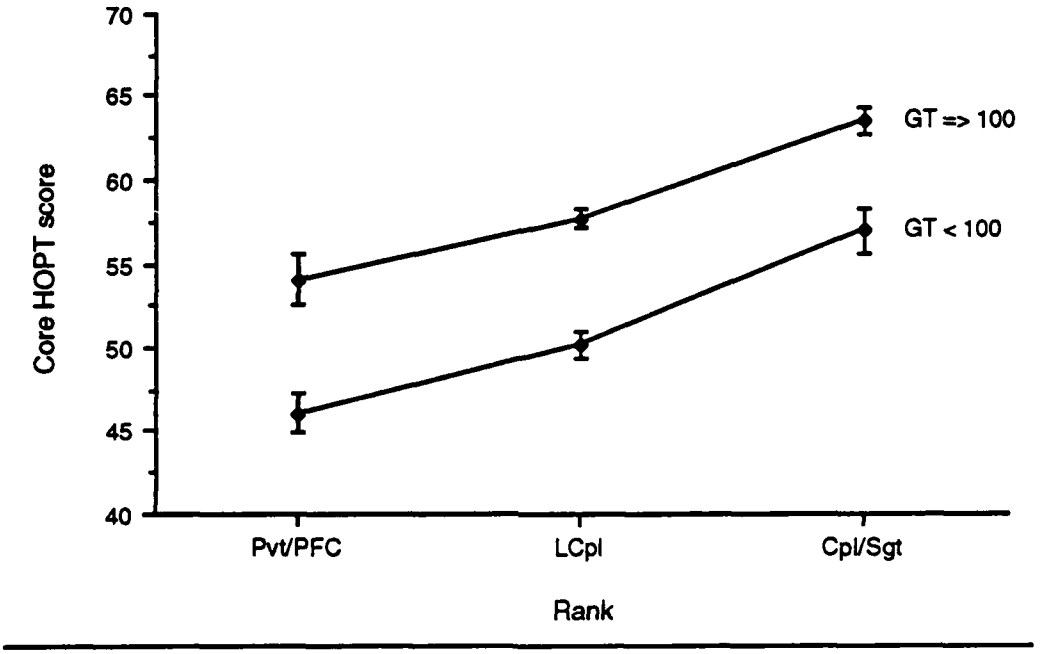

Figure 11. Mean core hands-on performance test by rank and aptitude level

Regardless of the aptitude group, higher ranks were able to perform significantly better than their subordinates. This result confirms that the Marine Corps promotion system is properly advancing its higher performers. Unlike the results for time in service, the proficiency trends for higher ranks did not stall. The lack of improvement at higher periods of time in service may be due to a significant number of low-performing and low-ranking personnel who simply have been enlisted for a long time (e.g., persons who have been demoted). Thus, rank is a more refined measure of experience that incorporates both quality of past performance and time in service.

As with time in service, the higher aptitude group was able to achieve significantly better hands-on proficiency levels than the low-aptitude group for each

rank category. In addition, some interesting tradeoffs exist between aptitude and rank, as seen in figure 11. First, comparable hands-on proficiency levels were achieved by high-aptitude LCpls and low-aptitude Cpl/Sgts. Second, Pvt/PFCs of high aptitude performed only slightly less than Cpl/Sgts of low aptitude. Finally, substantial proficiency differences were found between high-aptitude Pvt/PFCs and low-aptitude LCpls; the Pvt/PFCs performed significantly better. Thus, despite increasingly more difficult job requirements for higher ranks, high-aptitude personnel at lower ranks were better able to assimilate more complex concepts or to model their superiors' performance so as to significantly exceed the proficiency of higher ranking but lower aptitude Marines. Although the hands-on tests did not directly measure leadership or the ability to work in small units, stark differences were revealed in overall proficiency to perform critical infantry tasks relative to aptitude and rank. Such comparisons of proficiency by rank are useful in developing the case for high-quality recruits because it is both an experience and performance metric readily understood by and operationally meaningful to the Marine Corps.

Similar outcomes were noted in figure 12 for job knowledge comparisons by aptitude and rank. Significant performance differences between high- and low-aptitude groups on the JKTs existed for each rank category. Again, high-aptitude Pvt/PFCs performed as well as low-aptitude Cpl/Sgts.

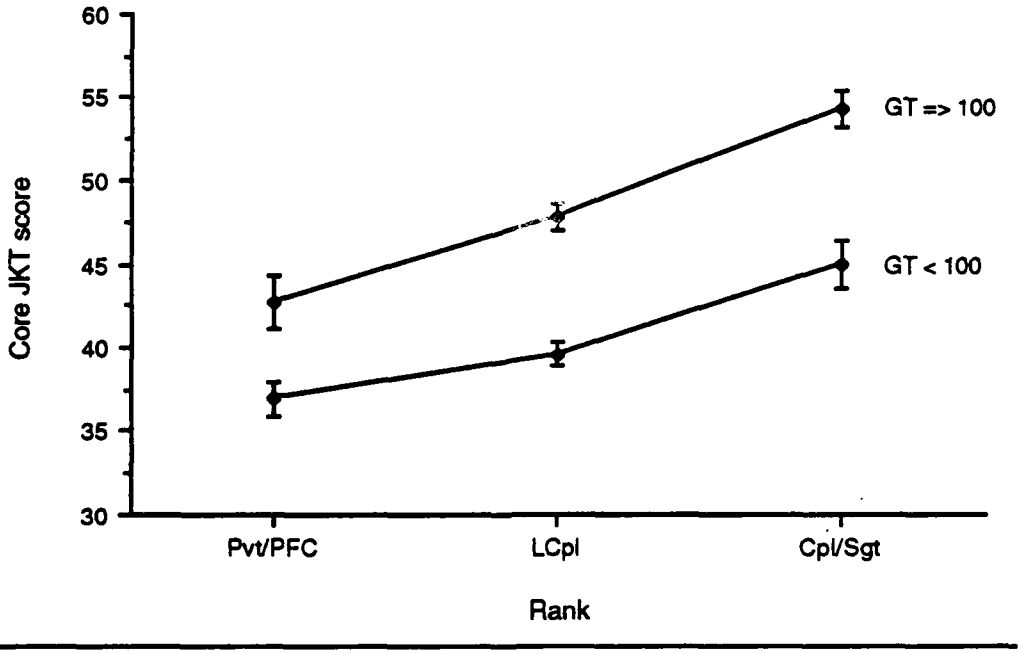

Figure 12. Mean core job knowledge test by rank and aptitude level

Recency of Task Performance

If a Marine has had recent opportunities to perform infantry tasks in a training setting or on an exercise, his ability to perform those tasks in a testing situation may be affected. Information regarding the recency of task performance was collected from each examinee so that past experiences could be considered in interpreting individual proficiency scores.

Before performing each task, the examinee was asked how recently he had performed the task: less than one week earlier, less than one month earlier, less than six months earlier, greater than six months earlier, or never (had received instruction only). Although this information was based on self-reports, the data were assumed to be reliable. The recency of task experience for each Marine was averaged to determine an overall measure of recency with infantry tasks.

Recency of task performance may have implications for examining the interaction between aptitude and frequency of training. Is training able to compensate for lower aptitude in the prediction of job performance? Using recency of task performance as a proxy for training, figure 13 shows the relationship between aptitude and recency of hands-on performance. Recency of performance had a positive impact on hands-on proficiency. Those persons with more recent task performance experience (within the past month) performed significantly better than those who had limited practice (experience greater than one month earlier) or who had instruction only and no performance experience. This finding was consistent for both high- and low-aptitude groups.

Figure 13 also reveals that the hands-on proficiency differences between high- and low-aptitude groups were not significant for the groups who had "instruction only" (although sample sizes for the two groups were small for this condition). Thus, for training settings in which individuals were merely exposed to the material and had no opportunities to perform the task, differences in hands-on proficiency were not significant for high- versus low-aptitude groups. Given "limited" opportunities to perform infantry tasks, however, the high-aptitude group showed significant gains in hands-on proficiency. This was not true for the low-aptitude group, which showed only slight improvement.

Several general implications concerning infantry training can be drawn from these findings. First, a certain standard of performance can be attained by the high-aptitude group with substantially less training or on-the-job experience than would be required by the low-aptitude group. Figure 13 shows that high-aptitude

persons need only limited experience in task performance to perform as well as low-aptitude persons with much more recent task exposure. Second, significantly higher proficiency levels can be achieved by a high-aptitude group for a fixed level of training. Figure 13 shows that, given only limited performance opportunities, the high-aptitude group was about 16 percent more proficient than the low-aptitude group. The tradeoffs between aptitude and recency of task performance demonstrate that recency of performance does not compensate for low initial aptitude.

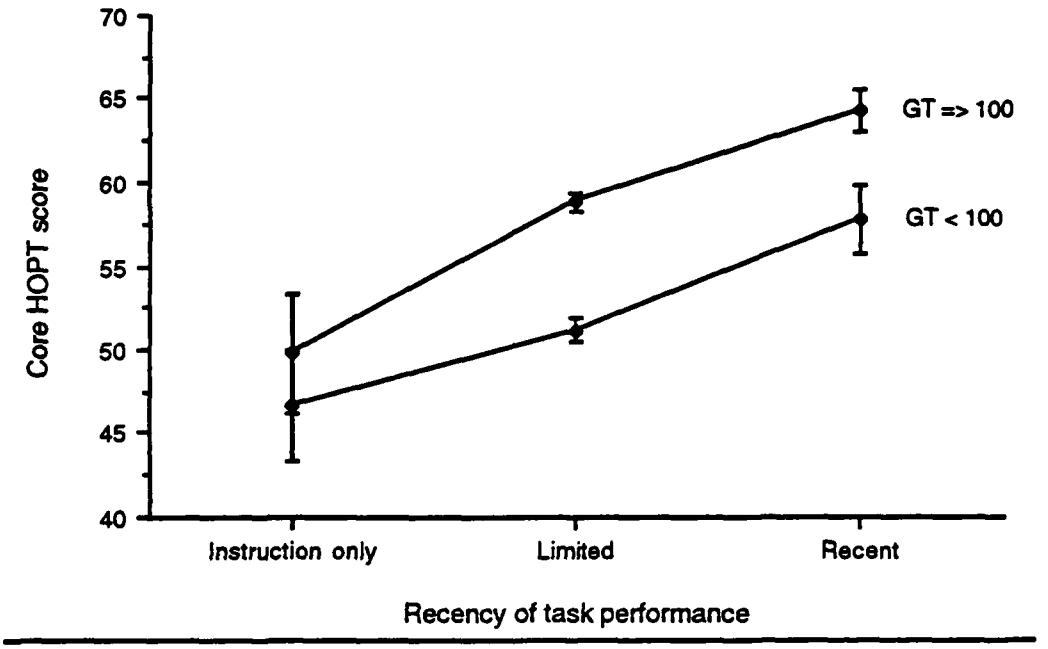

Figure 13. Mean core hands-on performance test by recency of task performance and aptitude level

The effect of recency of task performance had less impact on JKT proficiency than it did on HOPTs (see figure 14). Although the trend of JKT proficiency improved for the low-aptitude group that had more recent practice, the proficiency levels for each recency condition were not significantly different. The high-aptitude group showed significant improvement in JKT proficiency from the "instruction only" to "limited" conditions, but not from "limited" to "recent" exposure.

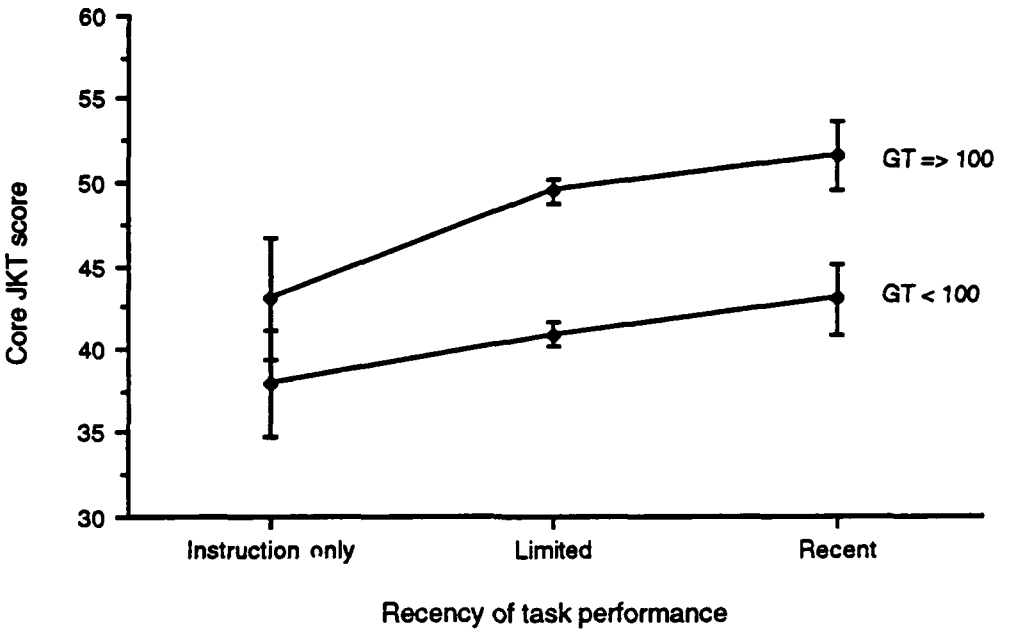

Figure 14. Mean core job knowledge test by recency of task performance and aptitude level

CONCLUSIONS

The Marine Corps selected the infantry occupational field for the initial JPM testing because it is the largest and often the most controversial for establishing aptitude requirements. Over 2,300 infantrymen in five MOSs were tested. The findings for each MOS were particularly noteworthy and provide the Marine Corps with a solid case for definitively illustrating the impact of recruit quality on infantry performance. ASVAB was found to be highly valid as a predictor of several different infantry performance measures: hands-on performance of critical infantry tasks, a written job knowledge test, and grades in training courses. Such results strongly dispute the common belief that anyone, regardless of his aptitude level, can function equally well as an infantryman. Aptitude was shown to be highly relevant to the successful performance of infantry job requirements.

The strong relationship between aptitude and performance also extended into the second term of service. Individuals with higher aptitudes were found to be the better second-term performers. This outcome has significant implications for the types of individuals the Marine Corps should recruit to be able to staff its leadership positions with qualified personnel in future years. In addition, these results indicate that the Marine Corps would have difficulty in quickly recovering from the

detrimental effect of low recruit quality. Declines in quality would have substantial long-term effects by reducing not only the level of first-term performance but also eventually affecting the caliber of enlisted leadership.

The validities against hands-on proficiency differed across the four Marine Corps aptitude composites. These findings justify the use of separate aptitude composites for recruit classification and imply that a general overall aptitude score is insufficient. As noted, however, the classification process may be improved by considering a composite more like the MM composite than the current GT composite. Across the five infantry MOSs, MM validities were typically 4 points better than the GT validities; such improvements are substantial (over 7 percent). The Marine Corps could benefit by revising its composite used for infantry classification.

Experience in the Marine Corps, as measured by time in service, rank, and recency of task performance, had a significant and positive influence on job proficiency. Despite such developmental effects, experience was not able to compensate for low aptitude in the prediction of proficiency. The initial gap in hands-on proficiency between high- and low-aptitude groups was so great that more than three years of service or the rank of at least corporal were required to offset the deficit. The recency of task performance ratings also showed that significantly more training or practice was required of lower aptitude personnel to achieve the same proficiency outcomes as those of higher aptitude personnel with less training or practice.

Although these validity results provide convincing evidence for the impact of aptitude on performance, the Marine Corps must address a secondary issue concerning what performance levels it requires from its infantrymen. Significant validities imply that "more aptitude is better," but it does not confront the issue of "how much performance is enough." This determination of required performance levels is an important ingredient in the Marine Corps' overall justification of its requirements for recruit quality. Establishing the aptitude/performance linkage is an initial and necessary, but not sufficient, condition in this process.

The Marine Corps JPM project succeeded in developing reliable, objective, and representative measures of infantry performance. The hands-on performance tests did not necessarily measure an individual's ability to lead, to operate in stressful or hostile environments, or to work effectively with others in small units. Also, the performance tests were not collected under combat-like conditions (the ultimate performance criterion); however, combat would only degrade the performance levels observed in the JPM project. Individuals currently not able to accomplish the

requisite infantry tasks in a nonthreatening environment probably would not be able to perform them under the chaotic conditions of war. The JPM project has empirically demonstrated a strong and persistent relationship between individuals' aptitudes and their future performance in the infantry occupational field and, ultimately, their ability to function more effectively in combat by being able to perform critical infantry skills.

REFERENCES

[1] A. K. Wigdor and B. F. Green, eds. *Assessing the Performance of Enlisted Personnel: Evaluation of a Joint-Service Research Project*. Committee on the Performance of Military Personnel, National Research Council. Washington, D.C.: National Academy Press, 1986

[2] MCO 1510.35A, *Individual Training Standards (ITS) System for Infantry, Occupational Field (OccFld) 03*, Training, Headquarters, U.S. Marine Corps, Jan 1986

[3] CNA Research Contribution 570, *Developing a Competency Scale for Hands-On Measures of Job Proficiency*, by Paul W. Mayberry, Dec 1987 (02057000)[1]

[4] CNA Research Memorandum 88-259, *Analysis of Data Quality for the Infantry Phase of the Marine Corps Job Performance Measurement Project*, by Paul W. Mayberry, Mar 1989 (27880259)

[5] Richard J. Shavelson et al. "Generalizability of Job Performance Measurements: Marine Corps Rifleman." *Military Psychology* 2: 129-144

[6] H. Gulliksen. *Theory of Mental Tests*. New York: Wiley, 1950

[7] CNA Report 102, *Validity of the Armed Services Vocational Aptitude Battery Forms 8, 9, and 10 With Applications to Forms 11, 12, 13, and 14*, by Milton H. Maier and Ann R. Truss, Feb 1985 (94010200)

[8] CNA Study 1160, *Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 With Applications to ASVAB Forms 8, 9, and 10*, by William H. Sims and Catherine M. Hiatt, Feb 1981 (07116000)

[9] CNA Research Contribution 537, *Supervisor Ratings Analysis*, by Catherine M. Hiatt, Feb 1986 (02053700)

[10] John E. Hunter, James J. Crosson, and David H. Friedman, *The Validity of the Armed Services Vocational Aptitude Battery (ASVAB) for Civilian and Military Job Performance*, U.S. Air Force Contract No. F416689-83-C-0025, Aug 1985

1. The number in parentheses is a CNA internal control number.

APPENDIX A

SAMPLE AND CORRECTED RELIABILITIES OF INFANTRY PERFORMANCE MEASURES

APPENDIX A

SAMPLE AND CORRECTED RELIABILITIES OF INFANTRY PERFORMANCE MEASURES

Similar to validity coefficients, estimates of reliability are also affected by range restriction due to the selection process. An estimate of the population reliability coefficient ($\hat{\rho}_{xx}$) can be computed as follows:

$$\hat{\rho}_{xx} = 1 - \frac{s_x^2}{\sigma_x^2}\left(1 - r_{xx}\right) \quad , \tag{A-1}$$

where, s_x^2 and σ_x^2 are the sample and population variances, respectively, and r_{xx} is the sample reliability. Equation A-1 assumes that the error variances are equal for both the sample and population. Given that the infantry performance measures have no ceiling or floor effects, this assumption should be satisfied.

Table A-1 provides the sample and population standard deviations for the hands-on and job knowledge tests. Both the sample and computed population reliabilities are also presented. Population standard deviations were obtained from the range correction program that accounts for selection effects from all ASVAB subtests.

Table A-1. *Sample and population standard deviations for hands-on performance tests and job knowledge tests*

	MOS				
Test	0311	0331	0341	0351	0369
HOPT					
Sample	8.97	7.92	8.76	6.69	9.61
Population	10.14	9.52	9.28	7.15	10.70
JKT					
Sample	9.06	9.41	9.32	7.82	9.24
Population	12.07	12.14	10.49	9.35	10.73

Table A-2. Sample reliability estimates for hands-on performance tests and job knowledge tests

	MOS				
Test	0311	0331	0341	0351	0369
HOPT					
Test-retest	.70				
Alpha coefficient	.87	.87	.88	.83	.86
Scorer agreement	.90	.90	.89	.90	.90
JKT					
Test-retest	.73				
Alpha coefficient	.89	.89	.90	.87	.90

Table A-3. Corrected reliability estimates for hands-on performance tests and job knowledge tests

	MOS				
Test	0311	0331	0341	0351	0369
HOPT					
Test-retest	.77				
Alpha coefficient	.90	.91	.89	.85	.89
Scorer agreement	.92	.93	.90	.91	.92
JKT					
Test-retest	.85				
Alpha coefficient	.94	.93	.92	.91	.93

APPENDIX B

SAMPLE AND CORRECTED CORRELATIONS OF INFANTRY PERFORMANCE AND APTITUDE MEASURES

APPENDIX B

SAMPLE AND CORRECTED CORRELATIONS OF INFANTRY PERFORMANCE AND APTITUDE MEASURES

Descriptive statistics for each infantry performance measure and the Marine Corps aptitude composites are presented in this appendix.

Table B-1. Descriptive statistics for performance and aptitude measures by MOS

	MOS									
	0311		0331		0341		0351		0369	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Performance measure										
HOPT	52.6	9.0	54.8	7.9	52.9	8.8	64.1	6.7	55.1	9.6
JKT	44.1	9.1	48.6	9.4	52.4	9.3	52.9	7.8	49.8	9.2
PRO	43.7	2.1	44.0	2.0	44.1	1.6	44.2	1.5	—[a]	—[a]
GPA A[b]	50.6	10.0 (523)								
GPA B[b]	51.1	10.0 (669)								
Aptitude measure										
GT	102.3	12.5	104.5	12.5	105.1	12.6	109.8	11.7	98.5	14.9
MM	102.1	14.5	106.1	13.8	105.4	14.3	110.3	13.4	100.3	15.9
EL	99.8	13.1	102.1	13.2	102.2	13.4	106.9	13.0	96.8	15.3
CL	101.3	10.5	101.1	10.8	102.9	11.8	105.1	10.4	95.6	14.5
AFQT	48.6	18.5	50.7	19.2	51.6	19.8	58.5	18.6	52.3	22.0
Sample size	940		271		253		277		387	

a. PRO marks are not given to 0369s.
b. Statistics for GPA include examinees from all MOSs. Sample size for each location is noted in parenthesis.

Table B-2. Sample and corrected validities for aptitude composites against hands-on test

MOS	GT	MM	EL	CL	AFQT
		Sample values			
Rifleman	.47	.52	.46	.34	.40
Machinegunner	.53	.57	.56	.43	.48
Mortarman	.43	.55	.45	.25	.33
Assaultman	.42	.45	.44	.32	.38
Unit leader	.48	.49	.47	.40	.32
		Corrected values			
Rifleman	.62	.66	.61	.51	.55
Machinegunner	.70	.72	.72	.62	.66
Mortarman	.48	.60	.49	.30	.38
Assaultman	.51	.54	.53	.41	.46
Unit leader	.60	.60	.59	.51	.46

Table B-3. Sample and corrected validities for aptitude composites against job knowledge test

MOS	GT	MM	EL	CL	AFQT
		Sample values			
Rifleman	.55	.54	.56	.50	.54
Machinegunner	.54	.55	.57	.51	.54
Mortarman	.53	.60	.54	.36	.46
Assaultman	.49	.46	.51	.40	.50
Unit leader	.63	.64	.61	.52	.52
		Corrected values			
Rifleman	.78	.73	.77	.74	.77
Machinegunner	.74	.72	.77	.73	.74
Mortarman	.65	.70	.65	.50	.57
Assaultman	.68	.64	.69	.61	.68
Unit leader	.73	.72	.71	.63	.64

Table B-4. Sample and corrected validities for aptitude composites against proficiency marks

MOS	GT	MM	EL	CL	AFQT
		Sample values			
Rifleman	.18	.18	.18	.17	.17
Machinegunner	.08	.08	.09	.13	.07
Mortarman	.04	.11	.07	.02	.01
Assaultman	.19	.16	.19	.13	.20
		Corrected values			
Rifleman	.32	.31	.32	.31	.31
Machinegunner	.22	.21	.24	.25	.22
Mortarman	.05	.11	.07	.01	.02
Assaultman	.11	.11	.13	.06	.11

Table B-5. Sample and corrected validities for aptitude composites against grade-point average

MOS	GT	MM	EL	CL	AFQT
		Sample values			
Base A	.45	.46	.43	.34	.41
Base B	.28	.27	.28	.23	.27
		Corrected values			
Base A	.61	.61	.60	.53	.58
Base B	.43	.42	.43	.38	.41

APPENDIX C

DESCRIPTIVE STATISTICS OF PERFORMANCE MEASURES FOR EXPERIENCE AND APTITUDE CATEGORIES

APPENDIX C

DESCRIPTIVE STATISTICS OF PERFORMANCE MEASURES FOR EXPERIENCE AND APTITUDE CATEGORIES

Descriptive statistics for the core infantry content that was administered to all MOSs for both the hands-on and job knowledge tests are presented in this appendix. The statistics are computed separately for high and low categories of the general technical (GT) aptitude composite These statistics were used to create figures 9 through 14 of the main text.

Table C-1. Descriptive statistics for time in service vs. aptitude categories

	GT category							
Time in service (months)	< 100				=> 100			
	N	Mean	SD	SE[a]	N	Mean	SD	SE[a]
	Descriptive statistics for core HOPT content							
1-12	215	46.1	7.4	0.5	198	54.6	8.2	0.6
13-24	194	49.8	7.8	0.6	372	58.1	7.8	0.4
25-36	141	53.1	9.1	0.8	221	60.7	8.3	0.6
37-72	154	53.8	9.7	0.8	274	61.3	8.0	0.5
	Descriptive statistics for core JKT content							
1-12	211	37.2	6.7	0.5	198	44.9	9.9	0.7
13-24	192	38.9	6.7	0.5	369	47.2	10.3	0.5
25-36	140	40.4	9.0	0.8	220	50.7	10.4	0.7
37-72	151	44.1	9.2	0.8	273	52.5	10.9	0.7

a. SE represents the standard error of the mean.

Table C-2. Descriptive statistics for rank vs. aptitude categories

	GT category							
	< 100				=> 100			
Rank	N	Mean	SD	SE[a]	N	Mean	SD	SE[a]
Descriptive statistics for core HOPT content								
Pvt/PFC	165	46.1	7.6	0.6	137	54.0	8.7	0.7
LCpl	420	50.2	8.2	0.4	617	57.7	7.5	0.3
Cpl/Sgt	210	57.0	9.7	0.7	377	63.5	8.0	0.4
Descriptive statistics for core JKT content								
Pvt/PFC	160	36.9	6.6	0.5	137	42.8	9.3	0.8
LCpl	417	39.6	7.5	0.4	613	47.8	9.9	0.4
Cpl/Sgt	207	44.9	10.5	0.7	376	54.3	11.0	0.6

a. SE represents the standard error of the mean.

Table C-3. Descriptive statistics for recency of task performance vs. aptitude categories

	GT category							
Recency of task performance	< 100				=> 100			
	N	Mean	SD	SE[a]	N	Mean	SD	SE[a]
Descriptive statistics for core HOPT content								
Instruct only	20	46.7	7.4	1.7	24	49.8	8.8	1.8
Limited	743	51.2	9.5	0.3	975	58.8	8.3	0.3
Recent	85	57.8	9.6	1.0	164	64.3	8.1	0.6
Descriptive statistics for core JKT content								
Instruct only	19	37.9	6.9	1.6	24	43.0	8.9	1.8
Limited	734	40.8	9.0	0.3	971	49.5	10.6	0.3
Recent	84	43.0	9.8	1.1	163	51.6	12.8	1.0

a. SE represents the standard error of the mean.